

Package ‘signal.hsmm’

September 29, 2014

Type Package

Title Predict Presence of Signal Peptides

Version 1.1

LazyData true

Date 2014-09-21

Description Predicts presence of signal peptides in eukaryotic protein using hidden semi-Markov models. The implemented algorithm can be accessed both from command line and GUI.

License GPL-3

URL <https://github.com/michbur/signal.hsmm>

Depends R (>= 3.0.0)

Imports seqinr, shiny

Suggests knitr, markdown, shinyAce

NeedsCompilation no

Repository CRAN

Author Michal Burdukiewicz [cre, aut],Piotr Sobczyk [aut]

Maintainer Michal Burdukiewicz <michalburdukiewicz@gmail.com>

Date/Publication 2014-09-29 18:30:18

R topics documented:

aaaggregation	2
benchmark_dat	3
degenerate	3
duration_viterbi	4
find_nhc	5
gui_signal.hsmm	5
hsmm_pred	6
hsmm_pred_list	6
is_protein	7
plot.hsmm_pred	7
pred2df	8
read_txt	8
read_uniprot	9
run_signal.hsmm	9
signal.hsmm	10
summary.hsmm_pred	11
summary.hsmm_pred_list	11

Index	13
--------------	-----------

aaaggregation	<i>Scheme for amino acid aggregation</i>
---------------	--

Description

Amino acids are grouped together in larger sets based on their physicochemical properties important in the recognition of signal peptide.

Usage

```
aaaggregation
```

Format

a list of length four containing one-letter name of amino acid

benchmark_dat	<i>Benchmark data set for signal.hsmm</i>
---------------	---

Description

This data set lists eukaryotic proteins added to UniProt database release 2014_07 between 2011 and 2014 (140 proteins with signal peptide and 280 randomly sampled proteins without signal peptide). All proteins were used in benchmark test comparing the performance of signal.hsmm and other signal peptide predictors.

Usage

```
benchmark_dat
```

Format

a list of [SeqFastaAA](#) objects. Slot sig contains the range of signal peptide (if any).

Source

[UniProt](#)

Examples

```
summary(benchmark_dat)
```

degenerate	<i>Degenerate protein sequence</i>
------------	------------------------------------

Description

'Degenerates' protein sequence by aggregating aminoacids to bigger groups.

Usage

```
degenerate(seq, aa_group)
```

Arguments

seq	character vector representing single aminoacid sequence.
aa_group	list of aminoacid groups to which sequence should be aggregated.

Value

a character vector.

Examples

```

sample_seq <- sample(seqin:::a()[-1], 30, replace = TRUE)
table(sample_seq)

#compared with aggregated sequence
deg_seq <- degenerate(sample_seq, aaaggregation)
table(deg_seq)

```

duration_viterbi *Compute most probable path with extended Viterbi algorithm*

Description

Viterbi algorithm for Hidden Markov Model with duration Valid only for special case on data

Usage

```
duration_viterbi(aa_sample, pipar, tpmpr, od, params)
```

Arguments

aa_sample	character vector representing single aminoacid sequence.
pipar	Probabilities of initial state in Markov Model.
tpmpr	Matrix with transition probabilities between states.
od	Matrix of response probabilities. Eg. od[1,2] is a probability of signal 2 in state 1.
params	Matrix of probability distribution for duration. Eg. params[10,2] is probability of duration of time 10 in state 2.

Value

A list of length four:

- path a vector of most probable path
- viterbi values of probability in all intermediate points,
- psi matrix that gives for every signal and state the previous state in viterbi path,
- duration matrix that gives for every signal and state gives the duration in that state on viterbi path.

Note

Currently has very restricted application to specific input All computations are on logarithms of probabilities

find_nhc	<i>Localize n-, h- and c-region in signal peptide</i>
----------	---

Description

Finds borders between distinct regions constituting signal peptides using a heuristic algorithm.

Usage

```
find_nhc(protein, signal = NULL)
```

Arguments

protein	a vector of amino acids or object of class SeqFastaAA
signal	range of signal peptide. If NULL, the attribute sig of protein will be used.

Value

a vector of length 4 containing positions of:

1. start of n-region,
2. start of h-region,
3. start of c-region,
4. cleavage site.

References

Henrik Nielsen, Anders Krogh (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. Proc. Sixth Int. Conf. on Intelligent Systems for Molecular Biology.

gui_signal.hsmm	<i>GUI for signal.hsmm</i>
-----------------	----------------------------

Description

A graphical user interface for predicting presence of signal peptides.

Usage

```
gui_signal.hsmm()
```

Value

null.

Note

Any ad-blocking software may be cause of malfunctions.

See Also

[run_signal.hsmm](#)

Examples

```
gui_signal.hsmm()
```

hsmm_pred	<i>hsmm_pred class</i>
-----------	------------------------

Description

A single prediction of `signal.hsmm`.

Details

Always a named list of five elements

1. `sp_probability` is a probability of signal peptide presence.
2. `sp_start` is a start of potential signal peptide (naively 1 aminoacid).
3. `sp_end` is a position of last amino acid of signal peptide.
4. `struc` is numeric vector representing predicted structure of input protein.
5. `prot` is character vector containing input sequence of amino acids.

See Also

[summary.hsmm_pred](#) [plot.hsmm_pred](#)

hsmm_pred_list	<i>hsmm_pred_list class</i>
----------------	-----------------------------

Description

A list of prediction(s) generated by `run_signal.hsmm` function.

Details

A named list. Each element belongs to the `hsmm_pred` class.

See Also

[summary.hsmm_pred_list](#), [pred2df](#)

is_protein

Protein test

Description

Checks if an object is a protein (contains letters from one-letter amino acid code).

Usage

```
is_protein(object)
```

Arguments

object character vector where each elements represent one amino acid.

Value

TRUE or FALSE.

plot.hsmm_pred

Plot single signal.hsmm prediction

Description

Plots objects of class `hsmm_pred`.

Usage

```
## S3 method for class 'hsmm_pred'  
plot(x, add_legend = TRUE, ...)
```

Arguments

x object of class `hsmm_pred`.
add_legend logical, if TRUE, legend is added to the plot.
... ignored.

Value

Nothing.

pred2df	<i>Convert list of signal.hsmm predictions</i>
---------	--

Description

Converts objects of class [hsmm_pred_list](#) to data frame.

Usage

```
pred2df(object)
```

Arguments

object of class [hsmm_pred_list](#).

Value

Data frame which columns contain respectively the probability of signal peptide presence as well as the start and the end of predicted signal peptide.

read_txt	<i>Read sequences from .txt file</i>
----------	--------------------------------------

Description

Read sequence data saved in text file.

Usage

```
read_txt(connection)
```

Arguments

connection a [connection](#) to the text (.txt) file.

Details

The input file should contain one or more amino acid sequences separated by empty rows.

Value

a list of sequences. Each element has class [SeqFastaAA](#). If connection contains no characters, function prompts warning and returns NULL.

read_uniprot	<i>Read data from UniProt database</i>
--------------	--

Description

Read data saved in UniProt original flat text format.

Usage

```
read_uniprot(connection, euk)
```

Arguments

connection	a connection to UniProt data in text format.
euk	logical value if data has an eukaryotic origin.

Value

a list of sequences. Each element has class [SeqFastaAA](#). Slot sig contains the range of signal peptide. Sequence with more than one cleavage site or atypical aminoacids are removed.

run_signal.hsmm	<i>Predict presence of signal peptide in protein</i>
-----------------	--

Description

Using the hidden semi-Markov model predict presence of signal peptide in eukaryotic proteins.

Usage

```
run_signal.hsmm(test_data)
```

Arguments

test_data	single protein sequence (character vector) or list of sequences. It may be an object of class SeqFastaAA .
-----------	--

Details

Function `signal.hsmm` returns respectively probability of presence of signal peptide, start of signal peptide and the probable cleavage site localization. If input consists of more than one sequence, result is a data.frame where each column contains above values for different proteins.

Value

An object of class `hsmm_pred_list`.

Note

Currently start of signal peptide is naively set as 1 amino acid.

See Also

[hsmm_pred_list](#) [hsmm_pred](#)

Examples

```
#run signal.hsmm on one sequence
x1 <- run_signal.hsmm(benchmark_dat[[1]])

#run signal.hsmm on one sequence, but input is a character vector
x2 <- run_signal.hsmm(c("m", "a", "g", "k", "e", "v", "i", "f",
"i", "m", "a", "l", "f", "i", "a", "v", "e", "s", "s", "p", "i",
"f", "s", "f", "d", "d", "l", "v", "c", "p", "s", "v", "t", "s",
"l", "r", "v", "n", "v", "e", "k", "n", "e", "c", "s", "t", "k",
"k", "d", "c", "g", "r", "n", "l", "c", "c", "e", "n", "q", "n",
"k", "i", "n", "v", "c", "v", "g", "g", "i", "m", "p", "l", "p",
"k", "p", "n", "l", "d", "v", "n", "n", "i", "g", "g", "a", "v",
"s", "e", "s", "v", "k", "q", "k", "r", "e", "t", "a", "e", "s",
"l"))

#run signal.hsmm on list of sequences
x3 <- run_signal.hsmm(benchmark_dat[1:3])
#see summary of results
summary(x3)
#print results as data frame
pred2df(x3)
#summary one result
summary(x3[[1]])
plot(x3[[1]])
```

signal.hsmm

signal.hsmm - prediction of signal peptides

Description

Implementing hidden semi-Markov model and a novel approach to sequence analysis, signal.hsmm is new, highly accurate signal peptide predictor for eukaryotic proteins.

Details

signal.hsmm predicts presence of signal peptides using the hidden semi-Markov models.

Examples

```

few_predictions <- run_signal.hsmm(benchmark_dat[1:3])
#see all predictions
pred2df(few_predictions)
#summary one prediction
summary(few_predictions[[1]])
#plot one prediction
plot(few_predictions[[1]])

#have fun with GUI

gui_signal.hsmm()

```

```
summary.hsmm_pred      Summarize single signal.hsmm prediction
```

Description

Summarizes objects of class `hsmm_pred`.

Usage

```
## S3 method for class 'hsmm_pred'
summary(object, ...)
```

Arguments

```
object      of class hsmm_pred.
...         ignored
```

Value

Nothing.

```
summary.hsmm_pred_list      Summarize list of signal.hsmm predictions
```

Description

Summarizes objects of class `hsmm_pred_list`.

Usage

```
## S3 method for class 'hsmm_pred_list'
summary(object, ...)
```

Arguments

object of class [hsmm_pred_list](#).
... ignored

Value

nothing.

Index

- *Topic **classif**
 - run_signal.hsmm, 9
- *Topic **datasets**
 - aaaggregation, 2
 - benchmark_dat, 3
- *Topic **hplot**
 - plot.hsmm_pred, 7
- *Topic **manip**
 - degenerate, 3
 - pred2df, 8
 - read_txt, 8
 - read_uniprot, 9
 - summary.hsmm_pred_list, 11
- *Topic **methods**
 - plot.hsmm_pred, 7
 - summary.hsmm_pred, 11
 - summary.hsmm_pred_list, 11
- *Topic **print**
 - summary.hsmm_pred, 11
 - summary.hsmm_pred_list, 11

aaaggregation, 2

benchmark_dat, 3

connection, 8, 9

degenerate, 3

duration_viterbi, 4

find_nhc, 5

gui_signal.hsmm, 5

hsmm_pred, 6, 6, 7, 10, 11

hsmm_pred_list, 6, 8, 10–12

is_protein, 7

plot.hsmm_pred, 6, 7

pred2df, 6, 8

read_txt, 8

read_uniprot, 9

run_signal.hsmm, 6, 9

SeqFastaAA, 3, 5, 8, 9

signal.hsmm, 10

signal.hsmm-package (signal.hsmm), 10

summary.hsmm_pred, 6, 11

summary.hsmm_pred_list, 6, 11